

Yi: Open Foundation Models by 01.AI

Year: 2024 | Citations: 747 | Authors: 01.AI Alex Young, Bei Chen, Chao Li, Chengen Huang, Ge Zhang

Abstract

We introduce the Yi model family, a series of language and multimodal models that demonstrate strong multi-dimensional capabilities. The Yi model family is based on 6B and 34B pretrained language models, then we extend them to chat models, 200K long context models, depth-upscaled models, and vision-language models. Our base models achieve strong performance on a wide range of benchmarks like MMLU, and our finetuned chat models deliver strong human preference rate on major evaluation platforms like AlpacaEval and Chatbot Arena. Building upon our scalable super-computing infrastructure and the classical transformer architecture, we attribute the performance of Yi models primarily to its data quality resulting from our data-engineering efforts. For pretraining, we construct 3.1 trillion tokens of English and Chinese corpora using a cascaded data deduplication and quality filtering pipeline. For finetuning, we polish a small scale (less than 10K) instruction dataset over multiple iterations such that every single instance has been verified directly by our machine learning engineers. For vision-language, we combine the chat language model with a vision transformer encoder and train the model to align visual representations to the semantic space of the language model. We further extend the context length to 200K through lightweight continual pretraining and demonstrate strong needle-in-a-haystack retrieval performance. We show that extending the depth of the pretrained checkpoint through continual pretraining further improves performance. We believe that given our current results, continuing to scale up model parameters using thoroughly optimized data will lead to even stronger frontier models.